Strategic Customer Behavior in a Queueing System with a Loss Subsystem *

Refael Hassin[1] and Ran I. Snitkovsky[2]

[1,2]Department of Statistics and Operations Research, Tel Aviv University

September 15, 2016

Abstract

Inspired by cognitive radio applications, we study a noncooperative multi-player game of individual rational customers in a queueing network composed of two queues with different regimes. The first one is a free shared FIFO queue with waiting time affected by congestion. Wishing to avoid congestion, customers may choose to turn to the second server that offers service with no delay. However, the endeavor in reaching the second server is costly, and can go unrewarded because requests are blocked when the server is busy. Still, blocked customers do not leave the system empty handed – they are rerouted to the shared queue. The decision and benefit of each customer depend on the choices of others, bringing about a competition of utility-maximization among them. After analyzing the queueing characteristics of the system we show, by properties of the cost function, that a unique symmetric Nash equilibrium exists. Comparing the equilibrium strategy with the socially optimal strategy we find that in some cases, contrary to intuition, more customers choose the loss system than is socially preferred.

1 Introduction

It is sometimes the case that customers choose between anticipating different queues of different type or discipline (see [4] §8.2 for a survey). In particular, we refer to situations that involve choosing between a blocking subsystem and a shared subsystem of an unlimited capacity. A prominent application motivating the analysis of such decision problems comes from the framework of *cognitive radio networks* (CRNs).

Cognitive networks, due to their nature, are often associated with variations of unobservable queueing systems, in which customers are not informed of the queue length upon arrival (as in [1], [3], [7] and [9]). The concept of the unobservable M/M/1 queue was first introduced by Littlechild [11] and by Edelson & Hildebrand [2], and is covered in detail by Hassin & Haviv [5]. In [1] and [9], Do et al. and Jagannathan et al. study a generalization of the unobservable M/M/1 queue that models customer behavior in a CRN with users of two different types. Corresponding to [1] and [9], an observable variation of a queue with server breakdowns (in which customers make their decisions based on the observed queue length) is studied by Li & Han in [10].

---

*This research was supported by the ISRAEL SCIENCE FOUNDATION (grant No. 355/15)

One type of networks which, according to a survey paper by Haykin [8], is of a particular interest is the sensing-based networks. In sensing-based CRNs, users are endowed with the ability of sensing the channel, i.e., listening to the spectrum and detecting its state. Using this ability, users can determine whether or not the channel is idle and thus allocate their transmit power while maximizing their own benefit. As the act of sensing can be expensive (in terms of energy consumption, time, etc.), some users might waive sensing and join an unlicensed, freely-shared channel. Habachi & Hayel [3] investigate the decision process of users choosing between sensing and not sensing in a system consisting of two classes of customers, primary and secondary. Upon arriving to the system in [3], each customer chooses between joining a shared M/M/1 or applying (i.e. "sensing") for a free server in an M/M/s/s loss system.

Another related application is of *last mile delivery services*, where customers choose between two service regimes: door-to-door delivery or self pickup. Driven by this problem, Hayel et al. [7] consider a decision problem of customers choosing between an unobservable M/M/s/s and an unobservable M/D/1 queue. Other fields of practice include cloud computing services and the like.

The authors of [3] and [7] assume that sensing customers rejected by the blocking system leave never to return. This approach simplifies the analysis as it permits the division of the system into two queuing subsystems with two independent Poisson arrival streams. However, a model that permits customers to leave the system without being served raises concerns regarding the utility of service; In [3], only costs are considered and what in some cases blocking becomes beneficial for the customers. If on the contrary one assumes a fixed service reward, then customers must be given the opportunity of balking from the system when they are forced to bear a negative expected return. A natural question therefore would be how is the decision process changed when allowing blocked customers to join the shared queue?

Presented in this paper is a model closely related to the one in [3] and [7], but the assumptions made on the consequences of rejection make a significant difference. We study a network consisting of two servers, one queue, and homogeneous customers. One of the two servers in this model is a loss system (M/M/1/1); customers trying to enter the loss system (for the sake of avoiding congestion) are exposed to the risk of rejection. In addition, the act of sensing as itself, even when ended with rejection, is costly. As opposed to previously mentioned models, here we assume that sensing customers, upon encountering rejection are redirected to the shared queue. Thus, the queue in the system not only populates customers who initially choose to join it but also absorbs the blocked sensing customers. This assumption induces dependencies between the state of the loss system and the inter-arrival times of customers to the queue. More formally, the queue within this model functions as a modification of an M/M/1 queue in which the rate of arrivals is subject to Poisson alternations. Such Heterogeneous Arrivals Queues, were analyzed by Yechiali & Naor [12], and the results lay the foundations to the queueing analysis in this paper. For the sake of keeping the model simple and reducing the number of parameters in the problem, we assume that service times in both servers are of identical rates. We shall mention that all the results to be presented hereafter can be easily extended to the case of different exponential rates for each server.

Our key results in this paper are two: First we show that there exits a unique equilibrium strategy, which in general, but not always, is different from the socially-optimal strategy. Second, depending on the system parameters, customers may sense the loss system more than what the social planner would desire. In other words, sometimes customers who wish to avoid congestion cause a decrease in social welfare. Consequently, the price of anarchy is not a unimodal function of the system utilization. This arises counter intuitively inasmuch as selfish behavior in most models always yields to excessive joining to the shared queue, due to negative joining externalities (see [2],

[5] and [11]). Yet, these cases which we call *over-sensing* are rare and their significance in serving any practical purpose is uncertain.

We begin by presenting the mathematical model in §2. In §3 we derive stability conditions for the system. Afterwards, in §4 and §5 we find the equilibrium and the socially-optimal strategies and compare them using the concept of *price of anarchy*. Finally, in §6 we discuss possible future ways of generalizing the model for further investigation.

2 System-model

We consider a system composed of two identical servers, S_Q and S_L (Q for queue, L for loss) and a single FCFS unobservable queue (see Fig.1). Each service duration is exponential with rate μ. Customers are identical, and arrive at the system following a Poisson process with rate Λ. Upon arrival, a customer chooses one of two options: pay a *sensing price* and try to attain service in S_L, an action we shall call *sensing*, or join the shared queue and wait until accepted to service in S_Q, which will be named *joining*. Customers sensing S_L when the server is idle are immediately accepted to service, whilst sensing S_L when busy, they get rejected by S_L and redirect to the shared queue to wait their turn to service by S_Q.

Denote by p the probability that a customer senses S_L, and denote by $(X(t), Y(t))$ the state of the system at time t. $X(t) \in \{0, 1, 2, \ldots\}$ expresses the number of customers in the queue including the one in service in S_Q. $Y(t) \in \{0, 1\}$ expresses the state of S_L, where $Y(t) = 0$ means that S_L is idle at time t. Thus, the system can be described as a bi-dimensional Markov process, as shown in Fig.2. Let $P_{i,j}$ be the stationary probability that $X(t) = i$ and $Y(t) = j$ at some arbitrary moment t. In order to simplify the notation, when dealing with stationary probabilities we omit the use of t and denote the state by (X, Y), and $P_{i,j} = \Pr(X = i, Y = j)$. The stationary probabilities are computed through the following set of equations:

$$\begin{aligned} &\Lambda P_{0,0} - \mu P_{1,0} - \mu P_{0,1} = 0 \,, \\ &(\mu + \Lambda) P_{0,1} - p\Lambda P_{0,0} - \mu P_{1,1} = 0 \,, \end{aligned} \tag{2.1}$$

$$\forall n \in \{1, 2, \ldots\} : \begin{cases} (\mu + \Lambda) P_{n,0} - (1-p)\Lambda P_{n-1,0} - \mu P_{n+1,0} - \mu P_{n,1} = 0 \,, \\ (2\mu + \Lambda) P_{n,1} - p\Lambda P_{n,0} - \Lambda P_{n-1,1} - \mu P_{n+1,1} = 0 \,, \end{cases} \tag{2.2}$$

These equations lead to the following relationship:

$$\forall n \in \{0, 1, 2, \ldots\} : \; P_{n+1,0} + P_{n+1,1} = \frac{\Lambda}{\mu}((1-p)P_{n,0} + P_{n,1}) \,. \tag{2.3}$$

By the assumption that each individual chooses randomly and independently whether or not to sense, we deduce that subsystem S_L can be considered as an M/M/1/1 queue (Erlang's Loss Model) with arrival rate $p\Lambda$ and service rate μ. Denote $\rho := {}^{\Lambda}/_{\mu}$, then the *loss-probability* of S_L, i.e., the probability that a customer sensing S_L will find it busy, is:

$$\Pr(Y = 1) = \sum_{i=0}^{\infty} P_{i,1} = \frac{p\Lambda}{\mu + p\Lambda} = \frac{p\rho}{1 + p\rho}, \tag{2.4}$$

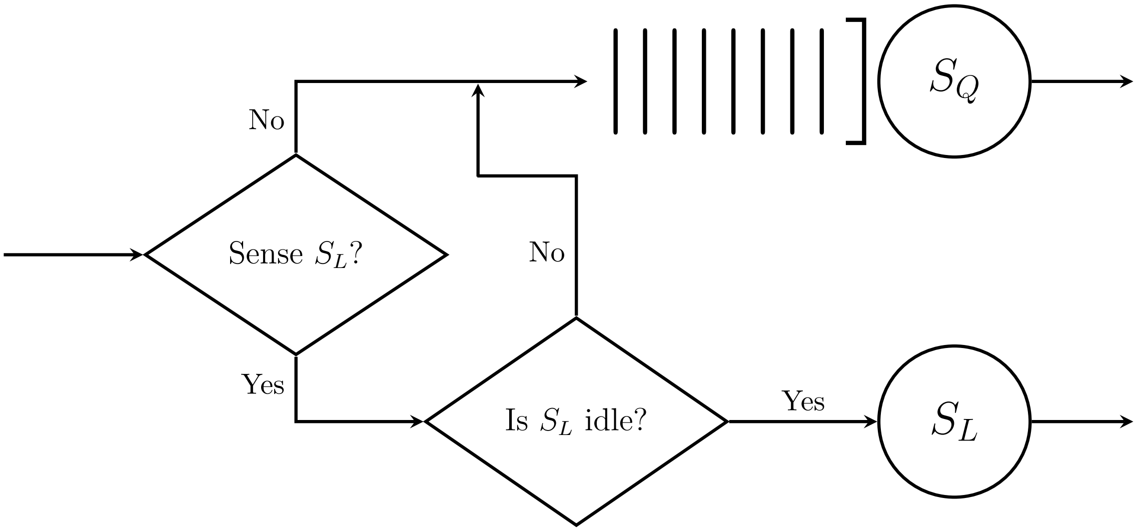


Figure 1: A flow chart of the system

and its complement:

$$\Pr(Y=0) = 1 - \Pr(Y=1) = \sum_{i=0}^{\infty} P_{i,0} = \frac{\mu}{\mu + p\Lambda} = \frac{1}{1+p\rho} \,. \tag{2.5}$$

Summing the equations in (2.3) over $n = 0, 1, 2, \dots$ and utilizing (2.4) and (2.5) we get:

$$P_{0,0} + P_{0,1} = 1 - \rho\left(1 - \frac{p}{1+p\rho}\right) \tag{2.6}$$

which indicates a linear relationship between $P_{0,0}$ and $P_{0,1}$.

Subtracting the second equation from the first one in (2.2) and combining with (2.3) we deduce the following formulae:

$$\forall n \in \{0,1,\dots\}: \begin{cases} P_{n+1,0} = (1-p)\rho P_{n,0} + p\rho \sum\limits_{i=0}^{n} P_{i,0} - \sum\limits_{i=0}^{n} P_{i,1} \,, \\ P_{n+1,1} = \rho P_{n,1} + \sum\limits_{i=0}^{n} P_{i,1} - p\rho \sum\limits_{i=0}^{n} P_{i,0} \,. \end{cases} \tag{2.7}$$

It can be seen from (2.7) that, for given p and ρ, every value $P_{n+1,k}$,$k \in \{0,1\}$, is expressed as a linear combination of the values $\{P_{0,0}, P_{1,0}, \dots, P_{n,0}, P_{0,1}, P_{1,1}, \dots, P_{n,1}\}$. Hence, with (2.6), it follows by induction that for each $n \in \{0,1,\dots\}$ and for each $k \in \{0,1\}$, $P_{n,k}$ is a linear function of $P_{0,0}$, as a countable sum of linear functions.

An accurate solution for (2.7) (including an analytic solution for $P_{0,0}$) is presented in a paper by Yechiali and Naor [12], but in an implicit form, since the solution makes use of Cardano's formula. The method in [12] can be utilized in evaluating the quantities $\mathrm{E}[X]$, $\mathrm{E}[X|Y=0]$ and $\mathrm{E}[X|Y=1]$ – the expected value, expected value given S_L is idle and the expected value given S_L is busy of the variable X, respectively.

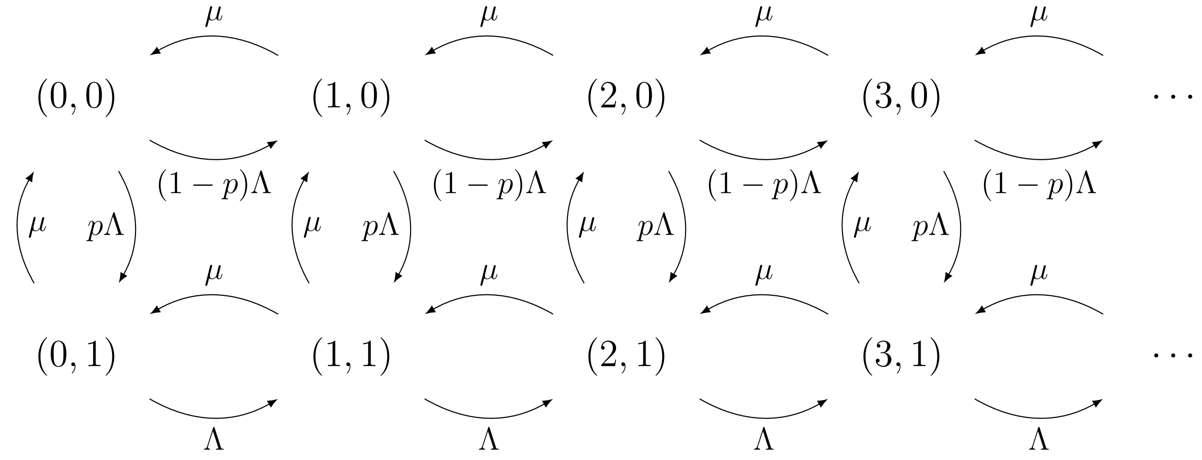


Figure 2: The Markov chain describing the transitions between states in the system

3 System-utilization

We now intend to find the maximum utilization of the system. The *effective-arrival-rate* to the shared queue, $\hat{\lambda}(p, \rho)$, consists of two streams of arrivals: customers who join the shared queue without sensing at all (whose proportion is $1 - p$) and customers who sense S_L and find it busy (whose proportion is $p \cdot \Pr(Y = 1)$). Therefore,

$$\hat{\lambda}(p, \rho) = (1 - p)\Lambda + p \cdot \Pr(Y = 1)\Lambda \,.$$

Using (2.4) and dividing by μ we get the *effective-utilization*, $\hat{\rho}(p, \rho)$:

$$\hat{\rho}(p, \rho) := \frac{1}{\mu}\hat{\lambda}(p, \rho) = \rho - p\rho + \frac{p\rho}{1 + p\rho}p\rho = \rho - \frac{1}{1 + p\rho}p\rho \,, \tag{3.1}$$

which is monotone decreasing in p for every $\rho \in (0, \infty)$.

Note, by (2.6), that $\hat{\rho}(p, \rho)$ is equal to $1 - (P_{0,0} + P_{0,1})$, for it represents the *busy fraction* of subsystem S_Q.

The stochastic process representing the stream of arrivals to the shared queue is not Poisson when $p > 0$. This is because the inter-arrival times to the queue are not i.i.d., and the arrival is higher when S_L is busy than when it is idle. This is a special case of queues with heterogeneous inter-arrival times presented in [12]. For the system to remain stable, as implied by [12], we assume $\hat{\rho}(p, \rho) < 1$.

Recall the *golden ratio*, $\varphi = 1/2 \cdot (1 + \sqrt{5}) \approx 1.618$. Then:

Proposition 3.1. *The system is stable if and only if* $\rho \in [0, \varphi)$ *and* $p \in (\underline{p}, 1]$, *where*

$$\underline{p} := \frac{\rho - 1}{\rho(2 - \rho)} \,.$$

Proof. A necessary and sufficient condition for stability is: $\hat{\rho}(p, \rho) < 1$. Substitute (3.1) and isolate p in the inequality to get an equivalent stability criterion:

$$p > \frac{\rho - 1}{\rho(2 - \rho)} = \underline{p}$$

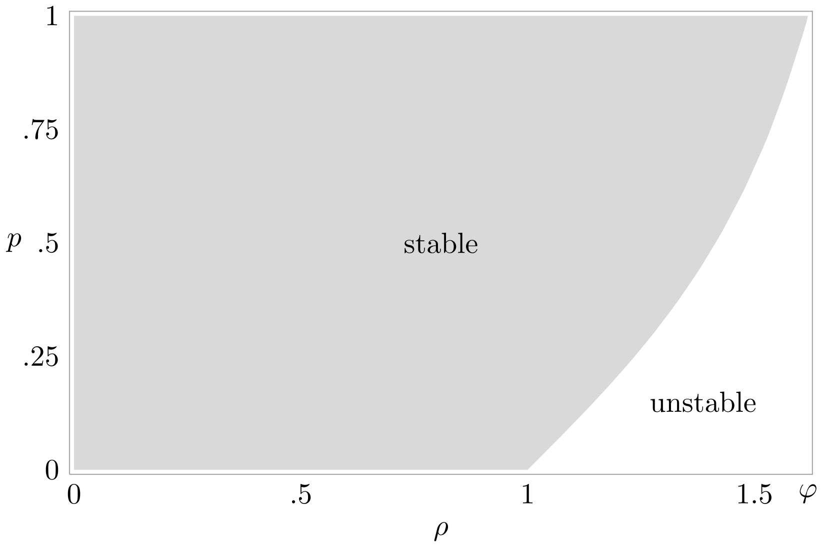


Figure 3: A colormap showing the p-ρ-plane divided into areas for which the system is stable and unstable. The shaded area in the graph shows values of p and ρ for which the system is stable, which are bounded by the curve $\underline{p}$.

from which it is clear that the system is stable for all $p \in [\underline{p}, 1]$.

Finding the maximum throughput we assume that $p = 1$, which means everyone senses S_L before joining the queue. This follows the monotonicity of $\hat{\rho}(p, \rho)$ as a function of p. Then, from (3.1), the stability criterion is:

$$\hat{\rho}(1, \rho) = \frac{\rho^2}{1+\rho} < 1 \quad \Leftrightarrow \quad \rho^2 - \rho - 1 < 0 \quad \Leftrightarrow \quad \rho < \frac{1+\sqrt{5}}{2} = \varphi$$

concluding that the system is stable for every $\rho \in [0, \varphi)$ and $p \in (\underline{p}, 1]$, otherwise it is overloaded. $\square$

Throughout the paper we assume that $\rho \in (0, \varphi)$ (when $\rho = 0$ the solution is trivial). Fig.3 shows the domain of possible pairs (ρ, p) such that the system is stable, which is bounded by the curve $\underline{p}$.

4 Equilibrium Strategy

In this section we discuss the (Nash) equilibrium strategies in the system. Let $c_w > 0$ be the cost per unit time of waiting in the shared queue, and $c_s > 0$ be the cost of sensing (incurred by each customer who chooses to sense, regardless of the result of this action). Since all customers are identical, and each one arriving at the system is eventually attained to service, neither the reward from service nor the time spent in service is relevant. Of course, the state of the system, including the length of the queue and the status of the servers, is unobservable as mentioned, and unknown to the customers at the time they make their decisions. We assume the system is not overloaded (i.e., $\rho \in (0, \varphi)$) and all customers act to reduce their expected cost to minimum. Denote by $C_S(p)$ and $C_N(p)$ the expected cost of an individual who chooses *sense* and *not sense*, respectively, given that the others' sensing probability is p. We have

$$\begin{cases} C_N(p) := \frac{c_w}{\mu}\mathrm{E}\left[X\right] \ , \\ C_S(p) := c_s + \Pr(Y=1)\cdot\frac{c_w}{\mu}\mathrm{E}\left[X \mid Y=1\right] \ . \end{cases} \tag{4.1}$$

We shall now show that there exists a unique equilibrium strategy. To this end, we first point out some properties of $\mathrm{E}\left[X \mid Y=0\right]$, of which we will make use in the proof:

Proposition 4.1. *The function* $\mathrm{E}\left[X \mid Y=0\right]$ *is continuous and monotone non-increasing in* p*.*

The proof is in the Appendix.

Denote $\gamma := {}^{c_w}/_{\mu c_s}$. The value of γ can be interpreted as the normalized cost of waiting a single service period, ${}^{c_w}/_{\mu}$, paid in currency with rate c_s. Similarly, ${}^{1}/_{\gamma}$ is the fixed normalized sensing cost.

Proposition 4.2. *For every* $\rho \in (0,\varphi)$*, and for every* $\gamma > 0$*, a unique equilibrium strategy* $p_e \in [0,1]$ *exists.*

Proof. $p \in [0,1]$ is an equilibrium strategy if no individual can benefit from choosing any alternative strategy. Recall C_S and C_N as defined in (4.1). Thus, the conditions for equilibrium are as follows:

- If for all $p \in [0,1]$ $C_N(p) \geq C_S(p)$ then $p=0$ is an equilibrium strategy.
- If for all $p \in [0,1]$ $C_N(p) \leq C_S(p)$ then $p=1$ is an equilibrium strategy.
- If for some $p \in (0,1)$ $C_N(p) = C_S(p)$ then p is an equilibrium strategy.

Other cases are infeasible considering that $C_N(p)$ and $C_S(p)$ are continuous.

Suppose first that there exists a value of p such that $C_N(p) = C_S(p)$. Note that:

$$\mathrm{E}\left[X\right] = \Pr(Y=1)\cdot\mathrm{E}\left[X \mid Y=1\right] + \Pr(Y=0)\cdot\mathrm{E}\left[X \mid Y=0\right] \ .$$

Substitute this expression in (4.1) and divide by c_s to get that $p_e = p \in (0,1)$ if and only if:

$$\gamma\Pr(Y=0)\cdot\mathrm{E}\left[X \mid Y=0\right] = 1 \, , \tag{4.2}$$

which, by integrating (2.5) is equivalent to

$$\gamma\mathrm{E}\left[X \mid Y=0\right] = 1 + p\rho \, . \tag{4.3}$$

Note that the function $1+p\rho$ is continuous and strictly monotone increasing in p, and therefore, using Proposition 4.1, we deduce that if there exists a solution p_e to (4.3) then it is unique, implying that p_e is a unique equilibrium strategy.

Suppose now that there is no p satisfying (4.3). Then if

$$\forall p \in (0,1) : \gamma\mathrm{E}\left[X \mid Y=0\right] > 1 + p\rho \, , \tag{4.4}$$

following Proposition 4.1 and the monotonicity of $1+p\rho$, we have that

$$\gamma\mathrm{E}\left[X \mid Y=0\right] \geq 1 + \rho \, , \tag{4.5}$$

or equivalently: $C_N(1) \geq C_S(1)$. This means that sensing is a dominant strategy for all individuals, and therefore $p_e = 1$ is a unique equilibrium strategy.

Finally, assume that

$$\forall p \in (0,1) : \gamma \mathrm{E}\left[X \mid Y = 0\right] < 1 + p\rho \,,$$

equivalently,

$$\gamma \mathrm{E}\left[X \mid Y = 0\right] \leq 1 \,, \tag{4.6}$$

implying that $C_N(0) \leq C_S(0)$, and $p_e = 0$, which is unique. □

This result is not as intuitive as it seems, since the function $C_S(p)$ is not necessarily monotone (as shown in Fig. 4a and 4b). This fact can be explained as follows: an increase in the proportion of customers that sense results in a decrease of the probability to attain service in S_L on one hand. It also decreases the expected queue length, and shortens the waiting time of the customers who failed to attain S_L on the other hand. Nevertheless, the equilibrium probability p_e is unique.

Proposition 4.3. *The pure strategy $p = 0$ is an equilibrium strategy (in other words $p_e = 0$) if and only if:*

$$\rho \leq \frac{1}{1+\gamma} \,.$$

Proof. Assuming that $\hat{\rho}(0,\rho) = \rho < 1$, from (4.6), $p = 0$ is an equilibrium strategy if and only if:

$$\gamma \mathrm{E}[X \mid Y = 0] \leq 1 \,.$$

Note that in this case the queue is an M/M/1 queue, and

$$\mathrm{E}[X \mid Y = 0] = \mathrm{E}[X] = \frac{\rho}{1-\rho} \,.$$

Thus, a necessary and sufficient condition for $p_e = 0$ is:

$$\gamma \frac{\rho}{1-\rho} \leq 1 \quad \Leftrightarrow \quad \rho \leq \frac{1}{1+\gamma} \,.$$

□

Proposition (4.3) provide a necessary and sufficient condition for $p_e = 0$. We would like to derive a similar condition determining when $p_e = 1$. Define $g(z)$ as

$$g(z) := (1-p)\rho^2 z^3 - (\rho^2 + (3-2p)\rho)z^2 + (2\rho + 2)z - 1 \,. \tag{4.7}$$

In their work, Yechiali and Naor [12] has shown that the polynomial $g(z)$ possesses a unique root, denoted z_0, in the interval $z \in (0,1)$, and that:

$$P_{0,0} = \frac{(1-\hat{\rho})z_0}{(1-z_0)(1-\rho z_0)} \,. \tag{4.8}$$

Define the partial generating function of the system for $Y(t) = 0$ as:

$$G_{Y=0}(z) := \sum_{m=0}^{\infty} z^m P_{0,m} \,, \quad |z| \leq 1 \,. \tag{4.9}$$

It has been proven by [12] that

$$G_{Y=0}(z) = \frac{1-\hat{\rho}+P_{0,0}(1-z)(\rho-1)}{g(z)} \,.$$

Note that

$$\mathrm{E}[X \mid Y=0] = \frac{1}{\Pr(Y(t)=0)} \cdot \frac{d}{dz} G_{Y=0}(z)\bigg|_{z=1} \,,$$

hence, denoting $g'(1) := ({}^{d}\!/_{dz}) g(z)|_{z=1}$,

$$\mathrm{E}[X \mid Y=0] = \frac{g(1)(1-\hat{\rho}-(\rho-1)P_{0,0}) - g'(1)(1-\hat{\rho})}{(g(1))^2 \cdot \Pr(Y(t)=0)} \,. \tag{4.10}$$

Proposition 4.4. *The pure strategy $p=1$ is an equilibrium strategy (in other words $p_e=1$) if and only if:*

$$\gamma \geq \frac{\theta^2+\theta^3-\theta^4}{1-\theta-\theta^2+\theta^3}$$

where $\theta := \sqrt{1+\rho}$.

Proof. Observing that for $p=1$ $g(z)$ becomes a polynomial of the second degree,

$$z_0 = \frac{1}{\rho}\left(1-\frac{1}{\sqrt{\rho+1}}\right) = \frac{1}{(1+\theta)\theta} \,.$$

With (4.8) we obtain

$$P_{0,0} = \frac{1+\theta-\theta^2}{\theta} \,.$$

Substituting this result in (4.10), after basic algebraic manipulations we get

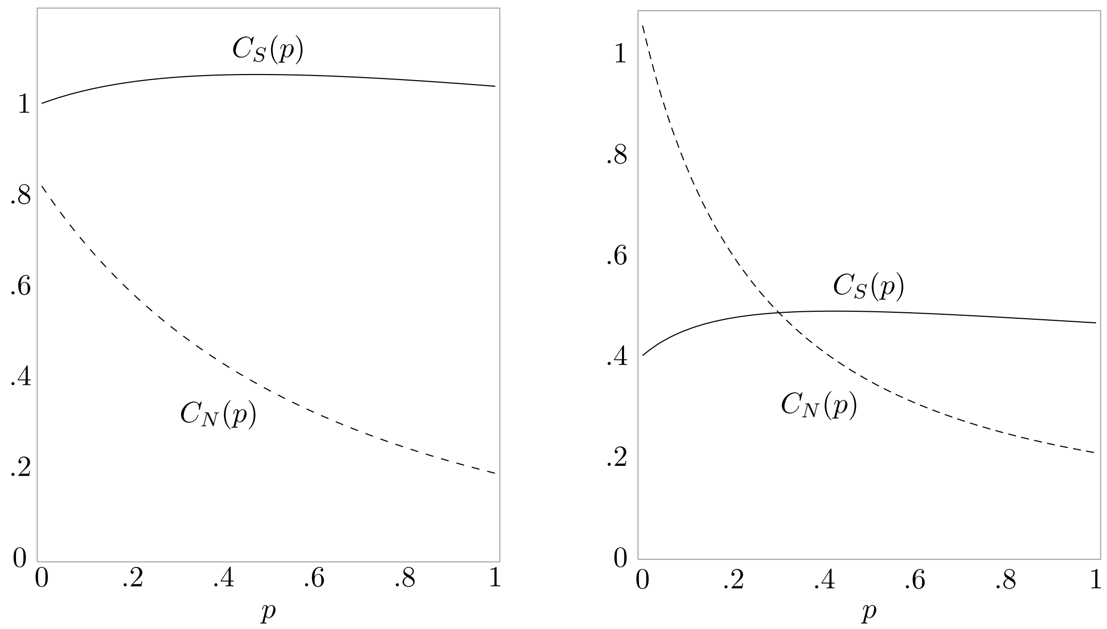
$$\mathrm{E}\left[X \mid Y=0\right] = 1 + \frac{\theta(2-\theta^2)}{1+\theta-\theta^2} \,.$$

Recall by (4.5) that $p_e=1$ if and only if $\gamma \mathrm{E}\left[X \mid Y=0\right] \geq 1+\rho$. We have then that $p_e=1$ if and only if

$$\gamma\left(1+\frac{\theta(2-\theta^2)}{1+\theta-\theta^2}\right) \geq \theta^2 \,. \tag{4.11}$$

Since $1+\theta-\theta^2>0$ for every $\rho \in (0,\varphi)$, we can multiply each sides of the inequality in (4.11) by $1+\theta-\theta^2$, and after rearrangement we get the desired result. □

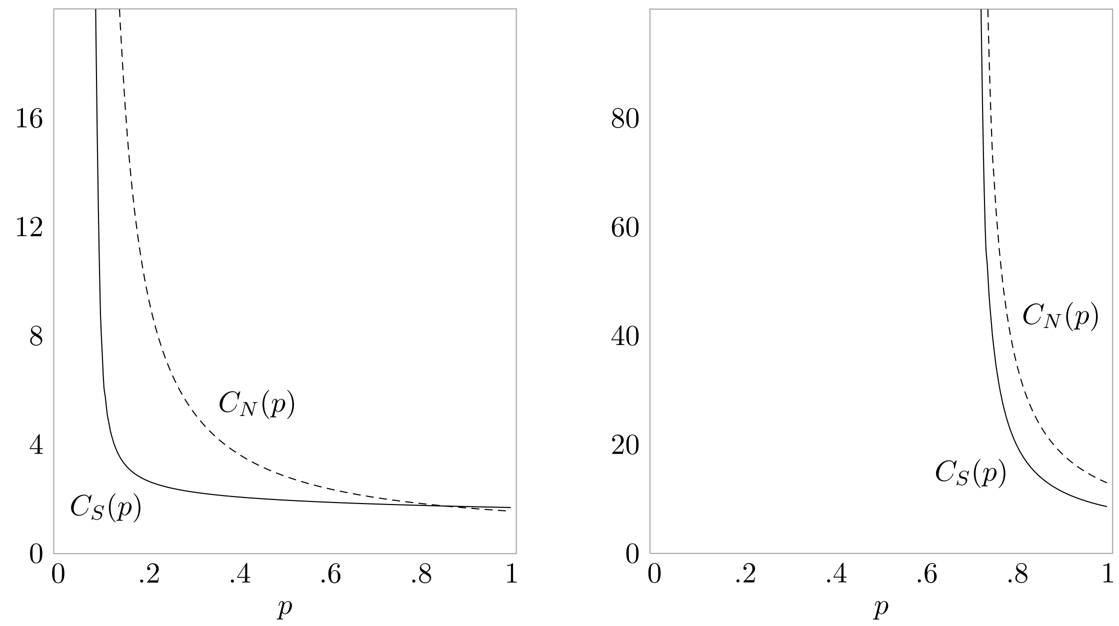
As a matter of fact, the technique of proving Proposition 4.4 is useful in formulating necessary and sufficient conditions not only for $p_e=1$ but also for every value $p_e \in (0,1)$. Put simply, for every set of parameters $\rho \in (0,\varphi)$ and $\gamma>0$ one can evaluate p_e analytically and without any computational difficulties. This is done by substitution of (4.10) in (4.3) and solving the equation for p (the solution depends on the root of the polynomial $g(z)$). In particular, when $\rho=1$, equation (4.10) yields $\mathrm{E}[X \mid Y=0] = {}^1\!/_p$, and applying the conditions for equilibrium it emerges that $p_e = \min\left\{{}^1\!/_2(\sqrt{1+4\gamma}-1), 1\right\}$. This result is reflected in Fig.6, Fig.7, and Fig.13a. Fig.8 depicts the value of p_e calculated for various values of ρ and γ.



(a) $\gamma = 1; \quad \rho = 0.45$

(b) $\gamma = 1; \quad \rho = 0.725$

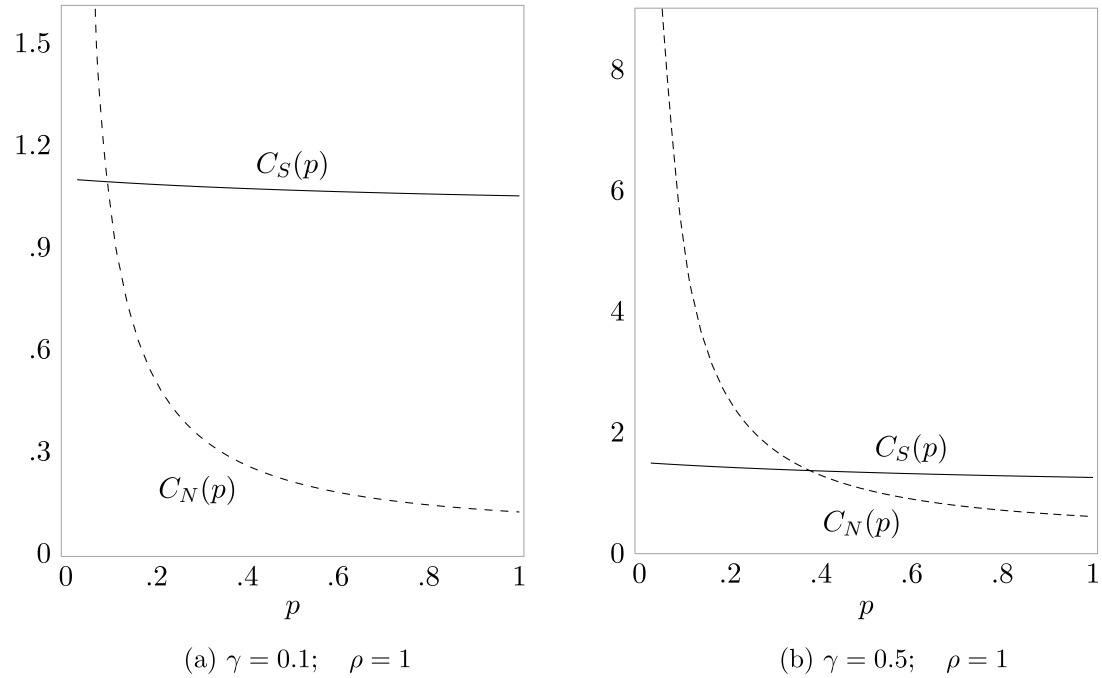
Figure 4: The expected costs of sensing (C_S) and not sensing (C_N) as functions of p cosidering $\gamma = 1$ and various values of ρ such that $\rho < 1$.



(a) $\gamma = 1; \quad \rho = 1.08$

(b) $\gamma = 1; \quad \rho = 1.515$

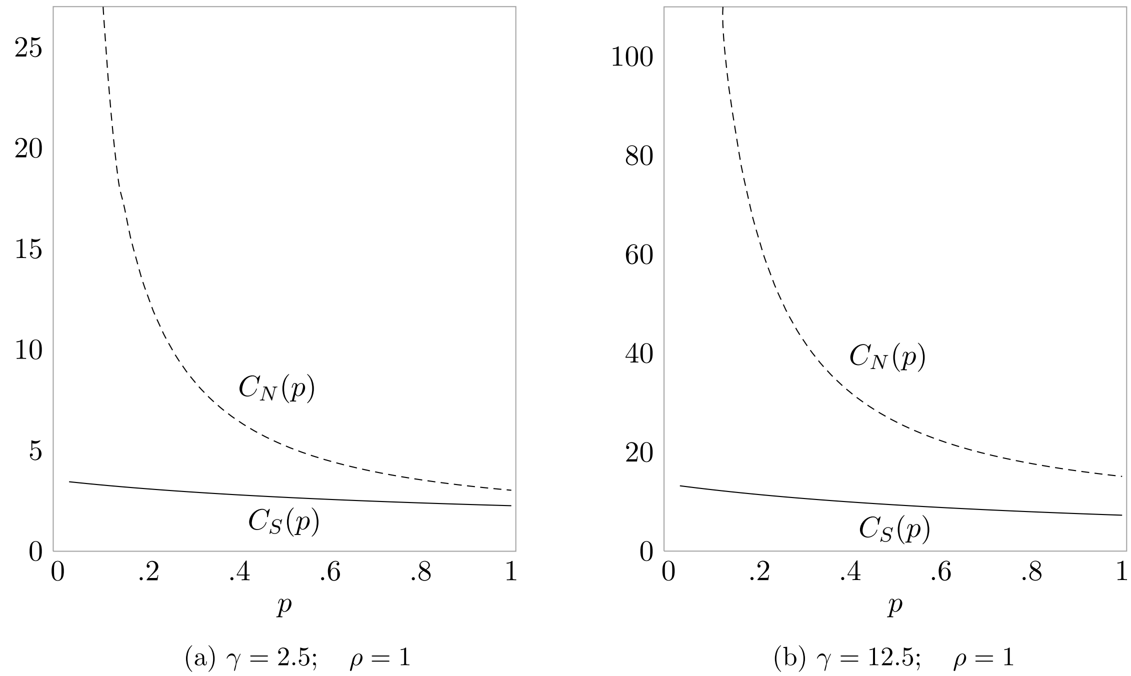
Figure 5: The expected costs of sensing (C_S) and not sensing (C_N) as functions of p cosidering $\gamma = 1$ and various values of ρ such that $\rho > 1$.



(a) $\gamma = 0.1; \quad \rho = 1$

(b) $\gamma = 0.5; \quad \rho = 1$

Figure 6: The expected costs of sensing (C_S) and not sensing (C_N) as functions of p cosidering $\rho = 1$ and various values of γ such that $\gamma < 1$.



(a) $\gamma = 2.5; \quad \rho = 1$

(b) $\gamma = 12.5; \quad \rho = 1$

Figure 7: The expected costs of sensing (C_S) and not sensing (C_N) as functions of p cosidering $\rho = 1$ and various values of γ such that $\gamma > 1$.

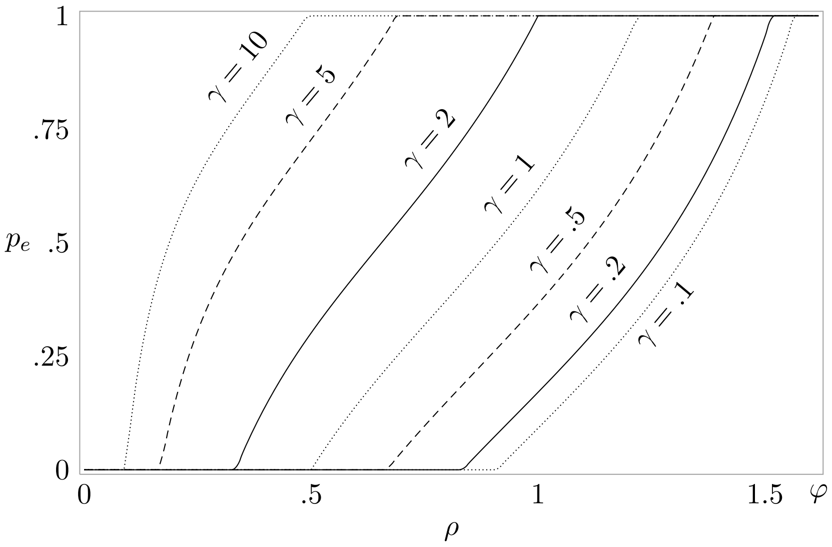


Figure 8: The equilibrium strategy p_e as a function of ρ for a various values of γ

5 Social Optimization

We now turn our attention to social optimization. The social objective function (normalized in c_w/μ), $C(p)$, is defined as

$$C(p) := \frac{\mu}{c_w}\left((1-p)C_N(p) + pC_S(p)\right) , \tag{5.1}$$

and represents the normalized expected cost of a customer when the probability of sensing is p. Combining (5.1) and (4.1) we obtain

$$C(p) = \frac{p}{\gamma} + (1-p)\cdot \mathrm{E}[X] + p\cdot \Pr(Y=1)\cdot \mathrm{E}[X \mid Y=1] . \tag{5.2}$$

Define $F(p)$ as follows:

$$F(p) := (1-p)\cdot \mathrm{E}[X] + p\cdot \Pr(Y=1)\cdot \mathrm{E}[X \mid Y=1] , \tag{5.3}$$

which implies that $C(p) = p/\gamma + F(p)$. $F(p)$ can be interpreted as the social cost function in a system of which $c_s = 0$ and $c_w/\mu = 1$ (that is to say sensing is free of charge and the expense of waiting for a single service period is 1 on average).

Proposition 5.1. *The function $F(p)$ is convex in p.*

The proof is in the Appendix.

Denote p^* the socially optimal strategy. Accordingly,

$$p^* := \underset{p\in[0,1]}{\arg\min}\, C(p) = \underset{p\in[0,1]}{\arg\min} \left\{\frac{p}{\gamma} + F(p)\right\} . \tag{5.4}$$

From Proposition 5.1 we immediately achieve that $C(p)$ itself is a convex function, thus, p^* is well defined.

Proposition 5.2.

(a) $p^* = 0$ *if and only if* $\rho \leq 1 - \sqrt{\gamma/1+\gamma}$.

(b) $p^* = 0 \;\Rightarrow\; p_e = 0$.

Proof.

(a) Suppose $p = 0$, then the system is an M/M/1 queue with arrival rate Λ. This is socially optimal if and only if society prefers to bear all the social cost imposed by an individual that chooses not to sense rather than pay the cost of sensing. We tag the "transparent" customer as the one assigning the lowest priority, that is to say preempted by every other customer. Note that if this tagged customer senses, he/she will find S_L idle and no other customer will preempt him/her because $p = 0$. It was shown in [5] that since the arrival and action of the transparent customer imposes no externalities on other customers, the optimal strategy for the transparent customer coincides with the socially optimal strategy. Thus, $p^* = 0$ if and only if the transparent customer prefers not to sense. As explained by [6], in an M/M/1 system, the queueing time of the transparent customer equals $1/\mu(1-\rho)^2 - 1/\mu$. This is the product of the expected queue length $(1/1-\rho)$ and the expected time it takes to reduce the queue size by one $(1/\mu(1-\rho))$ minus the expected service duration $(1/\mu)$. We deduce that $p^* = 0$ if and only if:

$$c_w\left(\frac{1}{\mu(1-\rho)^2} - \frac{1}{\mu}\right) \leq c_s \quad\Leftrightarrow\quad \gamma\left(\frac{1}{(1-\rho)^2} - 1\right) \leq 1 \quad\Leftrightarrow\quad \rho \leq 1 - \sqrt{\frac{\gamma}{1+\gamma}}\,.$$

(b) This result can be verified both algebraically and intuitively. Recall Proposition 4.3:

$$p_e = 0 \quad\Leftrightarrow\quad \gamma\frac{\rho}{1-\rho} > 1 \quad\Leftrightarrow\quad \rho \leq \frac{1}{1+\gamma}\,.$$

Note that:

$$1 - \sqrt{\frac{\gamma}{1+\gamma}} = \frac{1}{\sqrt{1+\gamma}}\left(\frac{1+\gamma-\sqrt{\gamma(1+\gamma)}}{\sqrt{1+\gamma}}\right) \leq \frac{1}{\sqrt{1+\gamma}}\left(\frac{1}{\sqrt{1+\gamma}}\right) = \frac{1}{1+\gamma}\,,$$

Where the second inequality follows from $\gamma \leq \sqrt{\gamma(1+\gamma)}$, concluding that if $\rho \leq 1 - \sqrt{\gamma/1+\gamma}$ then $\rho \leq 1/1+\gamma$.

We now give an intuitive explanation: Denote $W^0 := \mu^{-1}\cdot\rho/1-\rho$ the expected waiting time of a customer when $p = 0$ (as in the M/M/1 model). Suppose $p_e \neq 0$, then by Proposition 4.3, $\gamma\cdot\rho/1-\rho > 1$, which is equivalent to $c_w W^0 > c_s$. This means that an arbitrary customer can reduce both his/her own cost and the waiting cost of the others by sensing. Thus, $p^* \neq 0$, as claimed.

□

Proposition (5.2) implies that when $p_e = 0$ then $p^* \geq 0$. In fact, every $\gamma > 0$ satisfies

$$1 - \sqrt{\frac{\gamma}{1+\gamma}} < \frac{1}{1+\gamma}\,,$$

thus, taking $\rho \in [1 - \sqrt{\gamma/1+\gamma}, 1/1+\gamma]$ yields $p_e = 0 < p^*$. We shall now show that there exist values of ρ and γ such that $p^* < p_e$. In such cases we say that the system is in a situation of *over-sensing*. Although the value of p^* is computed in a numerical optimization technique, the existence of $p^* < p_e$ can be proved analytically. We begin by formulating a sufficient condition for $p^* < 1$:

Recall $F(p)$ as defined in (5.3).

Proposition 5.3. *If there exists a value* $h \in (0, 1)$ *such that*

$$\gamma \leq \frac{h}{F(1-h) - F(1)} \,, \tag{5.5}$$

then $p^* < 1$.

Proof. Denote $F'(1) := (d/dp)F(p)|_{p=1}$ and $C'(1) := (d/dp)C(p)|_{p=1}$. As discussed, $C(p)$ is convex, thus,

$$p^* < 1 \quad \Leftrightarrow \quad 0 < C'(1) = \frac{d}{dp}(\frac{p}{\gamma} + F(p))\bigg|_{p=1} \quad \Leftrightarrow \quad -\frac{1}{\gamma} < F'(1) \,.$$

Since $F(p)$ is convex we have that for all $h \in (0, 1)$:

$$\frac{F(1) - F(1-h)}{h} < F'(1) \,.$$

Suppose that $h \in (0, 1)$ satisfies (5.5), then

$$-\frac{1}{\gamma} \leq \frac{F(1) - F(1-h)}{h} < F'(1) \,,$$

and therefore $p^* < 1$. □

For a given fixed ρ and for h sufficiently small, define:

$$\gamma_1(\rho) := \frac{\theta^2 + \theta^3 - \theta^4}{1 - \theta - \theta^2 + \theta^3}; \quad \gamma_2(\rho) := \frac{h}{F(1-h) - F(1)} \,,$$

where $\theta = \sqrt{1+\rho}$. To this fixed value of ρ, the relation between these two quantities, $\gamma_1(\rho)$ and $\gamma_2(\rho)$ indicates whether there exists a value of γ such that $p^* < p_e$. Proposition 4.4 states that if $\gamma_1(\rho) \leq \gamma$ then $p_e = 1$. In addition, by Proposition 5.3, if $\gamma < \gamma_2(\rho)$ then $p^* < 1$. We deduce therefore that if $\gamma_1(\rho) < \gamma_2(\rho)$ then for all $\gamma \in [\gamma_1(\rho), \gamma_2(\rho))$, $p^* < p_e = 1$. This remarkable phenomenon can be spotted in Fig.11a, Fig.11b, Fig.12a and Fig.12b. In particular, taking $\rho = 0.6$ and $h = 0.01$ satisfies $\gamma_1(\rho) < \gamma_2(\rho)$ as demonstrated in Fig.12b. Binding the aforementioned result with Proposition 5.2, we conclude that in some rare cases, $p_e = p^* \notin \{0, 1\}$, namely p_e is an efficient non-trivial mixed equilibrium strategy. Cases as such can be spotted in Fig.14, which shows the two-dimensional ρ-γ-plane painted in different colors with respect to each of the three cases: $p^* > p_e$, $p^* = p_e$ and $p^* < p_e$.

In case $p^* < p_e$ then $1 - p_e < 1 - p^*$, and in order to reach optimal sensing rate customers should be incentivized to join the shared queue without sensing. We find this an unusual result, since in many models, and in Edelson & Hildebrand's fundamental model in particular (see [2]), the proportion of customers joining the queue in equilibrium is never less than the socially optimal proportion. The explanation is as follows: Say two successive customers arrive at the system one after another. Whereas in most models it is beneficial for the second customer that the first one will decide not to join the queue, in this model it is not clear whether the second one prefers that the first will sense or join.

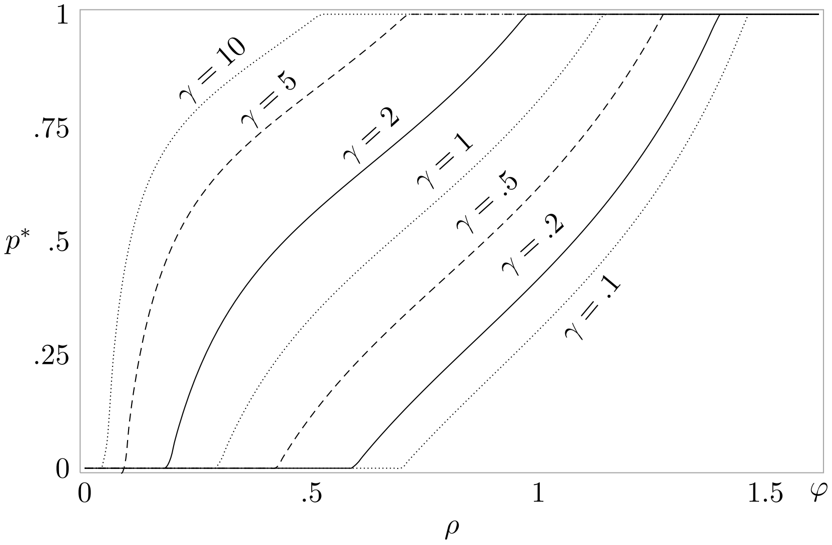
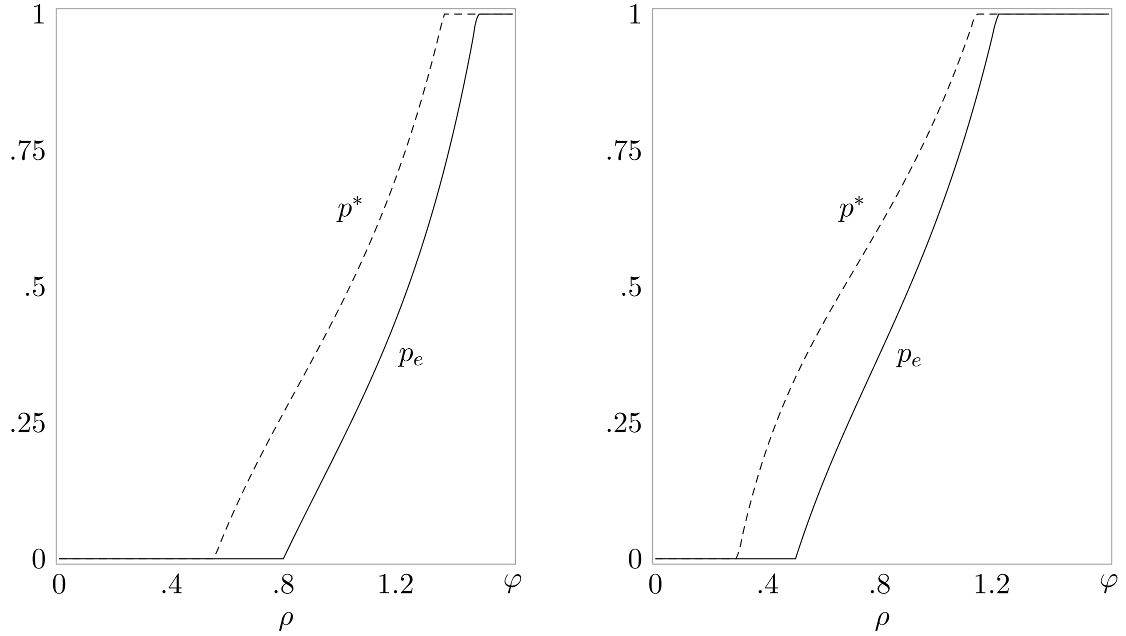


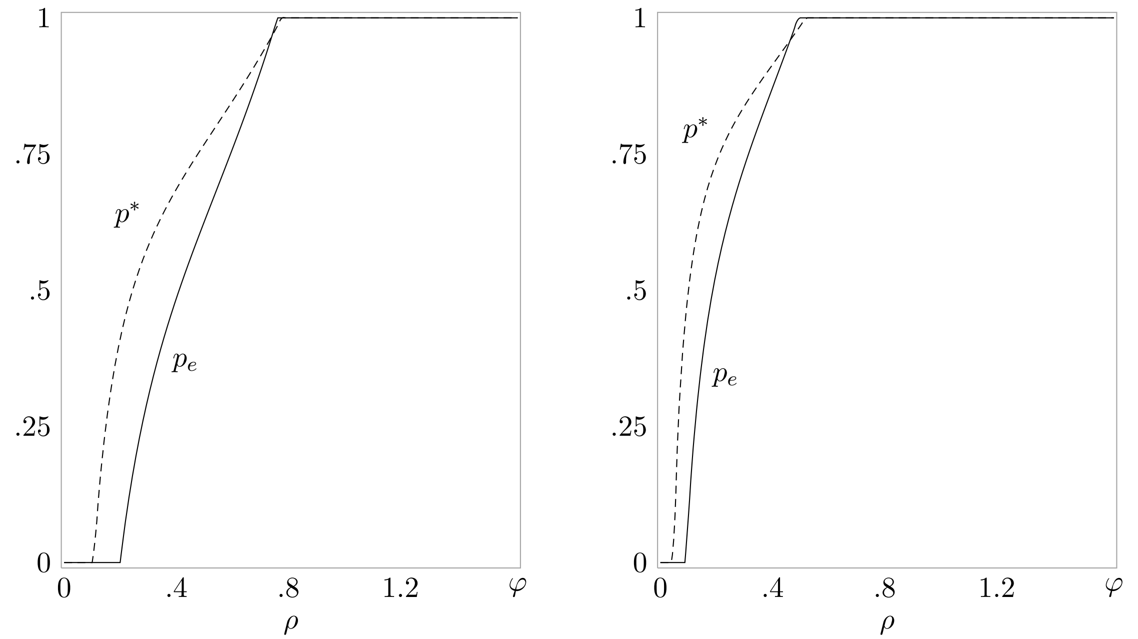
Figure 9: The socially optimal strategy p^* as a function of ρ for a various values of γ



(a) $\gamma = 0.25$

(b) $\gamma = 1$

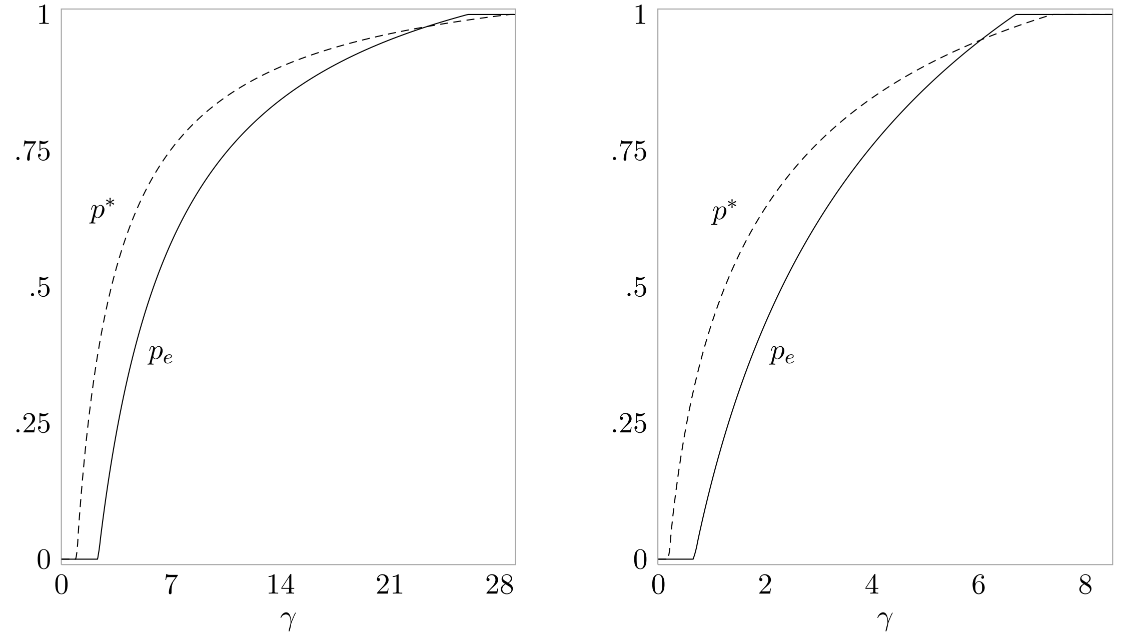
Figure 10: The socially optimal strategy (p^*) and the equilibrium strategy (p_e) as a function of ρ cosidering various values values of γ such that $\gamma \leq 1$.



(a) $\gamma = 4$

(b) $\gamma = 10$

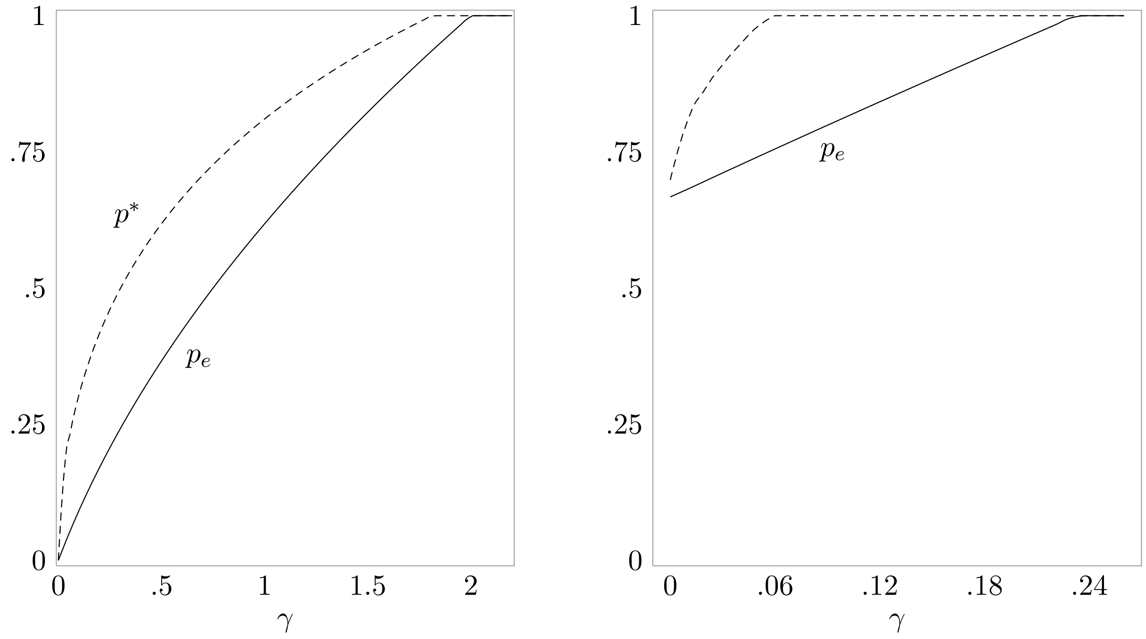
Figure 11: The socially optimal strategy (p^*) and the equilibrium strategy (p_e) as a function of ρ cosidering various values values of γ such that $\gamma > 1$.



(a) $\rho = 0.3$

(b) $\rho = 0.6$

Figure 12: The socially optimal strategy (p^*) and the equilibrium strategy (p_e) as a function of γ cosidering various values values of ρ such that $\rho \geq 1$.



(a) $\rho = 1$ (b) $\rho = 1.5$

Figure 13: The socially optimal strategy (p^*) and the equilibrium strategy (p_e) as a function of γ cosidering various values values of ρ such that $\rho \geq 1$.

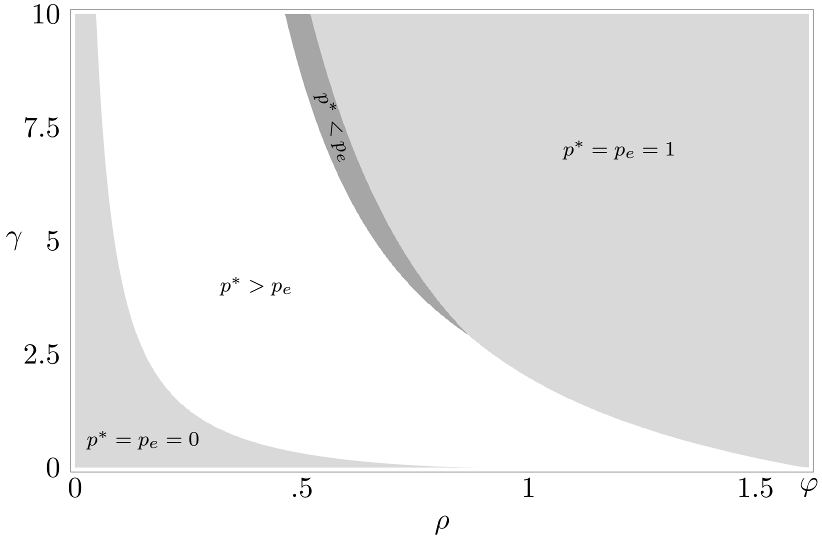


Figure 14: A color map of the two-dimensional ρ-γ-plane showing the relation between p^* and p_e. For each pair (ρ, γ), white indicates that $p^* > p_e$, gray indicates that $p^* = p_e$ (in particular where they are both equal to 0 or 1) and black indicates that $p^* < p_e$.

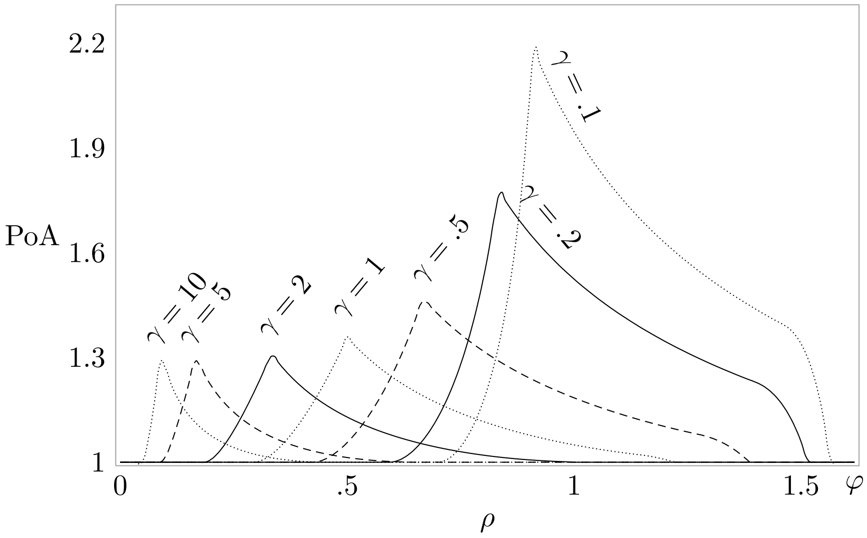


Figure 15: The Price of Anarchy (PoA) as a function of ρ for a various values of γ

5.1 Price of Anarchy

The existence of ρ and h such that $\gamma_1(\rho) < \gamma_2(\rho)$ implies that rational customers, under specific conditions, may over-sense S_L. This situation, while possible, may not be a matter of significant importance in practice, as the expected cost of customers in equilibrium is only slightly worse than that in social optimum under the circumstances of over-sensing. We reach this conclusion based upon numerical analysis of the *price of anarchy* (PoA), which is defined as follows:

$$\text{PoA}(\rho, \gamma) := \frac{C(p_e)}{C(p^*)} \,.$$

Empirical examination shows that the behavior of PoA as a function of ρ conforms with the results in Proposition 5.2 and Proposition 4.3 (see Fig.15). For a fixed value of γ:

- If $\rho \leq 1 - \sqrt{\frac{\gamma}{1+\gamma}}$ then $p_e = p^* = 0$ and PoA $= 1$.
- PoA (as a function of ρ) attains its maximum (non-smooth) point at $\rho = \frac{1}{1+\gamma}$.

where the first and the second properties follow from Proposition 5.2 and Proposition 4.3, respectively.

For a fixed given γ, we define ρ_1 such that $\gamma_1(\rho_1) = \gamma$. If $\gamma_1(\rho_1) < \gamma_2(\rho_1)$, then taking $\rho = \rho_1$ and γ to be the system parameters implies that $p^* < p_e = 1$. For this fixed value of γ, we notice that PoA is not a unimodal function of ρ, and attains a local maximum (non-smooth) point at ρ_1 in addition to its maximum point at ${}^1/_{1+\gamma}$. This can be observed in Fig.16, where $\gamma = 10$. Note that in the case where $\gamma = 10$ (see Fig.11b) then $\rho_1 \approx 0.49$, it holds that

$$\gamma_1(\rho_1) \approx 10.07 < \lim_{h \to 0} \gamma_2(\rho_1) \approx 11.23$$

and PoA$(0.49, 10) = 1.0005$, which means that $C(p_e)$ is marginally greater than $C(p^*)$.

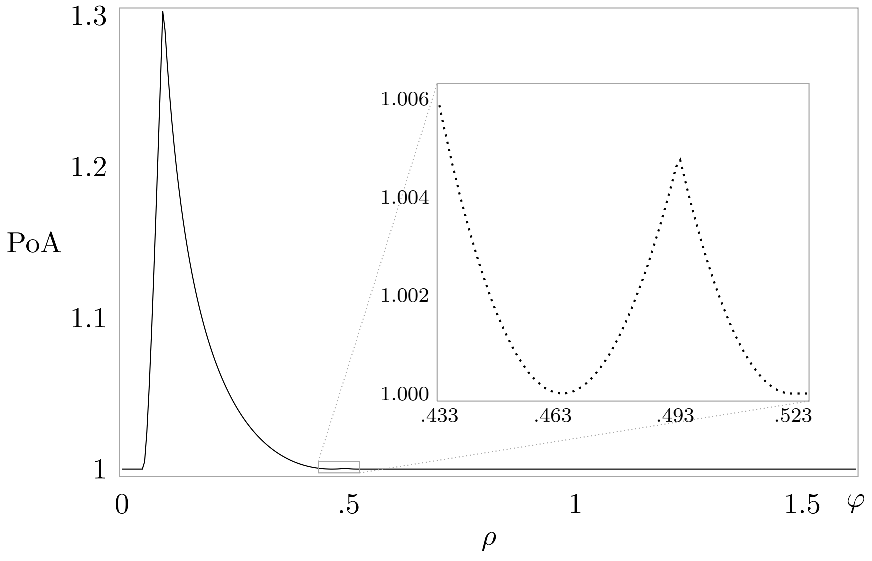


Figure 16: The Price of Anarchy (PoA) as a function of ρ for $\gamma = 10$. This example corresponds with Fig.11a, where $p^* = p_e$ (i.e., PoA = 1) at $\rho \approx .465$, for which the equilibrium strategy point is efficient and non-trivial. As shown, PoA is not unimodal in this case.

6 Concluding Remarks

This work deals with customer behavior in a queueing model inspired by the framework of CRNs. From the queueing analysis aspect, the main challenge is handling the dependency between the arrival rate to the queue and the customers' actions. In this specific model, customers' actions induce a queueing system in a random environment, in which the arrival stream is not a Poisson process.

We manage to show that the equilibrium strategy in this multi-player game is unique, and can be either greater, smaller or equal to the socially optimal strategy. This results prevail when extending the model to different service rates of S_Q and S_L, because the proof depends mainly on the fact that the service duration in both servers is memoryless and on the fact that the loss-probability of S_L is monotone increasing in p. Only slight changes are needed in order to apply the same solution for a network with different service rates, and the differences are primarily technical.

By proving properties (monotonicity and convexity) of the conditionally expected queue length and expected cost (which are functions of the sensing probability) we show the uniqueness of the equilibrium strategy, and compare it with the socially optimal one. However, the basic tools of calculus are of no use as these functions are given in an implicit form. In this paper, specifically in the proof of Proposition 4.1 and Proposition 5.1, we present a construction of an appropriate coupling argument that appear to be essential for the analysis of the model. Thus, we solve the problem of finding the equilibrium strategy analytically.

Further investigation and discussion is needed in order to apply the results in this work on models where there can be more than one server to sense. Extending the number of servers in a way that S_L becomes an M/M/s/s queue changes the state space of the system such that the solution of Yechiali & Naor [12] is no longer suitable. An accurate description of the decision problem in that case should regard the sensing methodology: how many servers are sensed, how they are sampled among the collection of servers etc. If we consider a case where each sensing customer is blocked if and only if all the loss system servers are all busy, then it can be shown in similar methods to those presented here, that the equilibrium strategy is unique.

7 Acknowledgments

The authors express their appreciation to Moshe Haviv, Liron Ravner and Uri Yechiali for many fruitful discussions.

8 Appendix

8.1 Proof of Proposition 4.1

Proof. Proving the continuity of $\mathrm{E}\left[X \mid Y=0\right]$ is immediate as, referring to (2.5), (2.6) and (2.7),

$$\mathrm{E}\left[X \mid Y=0\right] = \frac{1}{\Pr(Y=0)} \sum_{i=0}^{\infty} i P_{i,0} = (1+p\rho) \sum_{i=0}^{\infty} i P_{i,0} \ ,$$

which is a countable sum of continuous functions in $p \in [0,1]$.

In order to prove monotonicity, we examine the transitions between states in two systems, $\Omega := \{S_Q, S_L\}$ and $\Omega' := \{S'_Q, S'_L\}$, under the same sequence of events. Systems Ω and Ω' are identical except that in Ω the sensing probability is p, as opposed to Ω', where the sensing probability is p', and $p < p'$.

Denote by $(X(t), Y(t))$ and $(X'(t), Y'(t))$ the states at time t of systems Ω and Ω' respectively. Since for all $p \in [0,1]$ the state $(0,0)$ in the Markov chain is recurrent, we can assume w.l.o.g. that at time $t=0$,

$$(X(0), Y(0)) = (X'(0), Y'(0)) = (0,0) \ .$$

Showing that

$$\forall t \in [0,\infty) : \mathrm{E}[X'(t) \mid Y'(t)=0] \le \mathrm{E}[X(t) \mid Y(t)=0] \tag{8.1}$$

will complete the proof.

We begin by attaching three variables to each customer: Let $\{(T_i, \tau_i, u_i)\}_{i\in\mathbb{N}}$ represent the *arrival time*, *service duration* and *sensing aversion* of the i-th customer respectively, and for all $i \in \mathbb{N}$:

$$T_{i+1} - T_i \sim \exp(\Lambda) \ , \quad \tau_i \sim \exp(\mu) \ , \quad u_i \sim \mathrm{U}[0,1] \ . \tag{8.2}$$

Each customer arrives to both systems simultaneously. Customer i arriving at Ω senses if and only if $u_i \le p$ while arriving at Ω' he/she senses if and only if $u_i \le p'$. This ensures that the sensing probability is p in system Ω and p' in system Ω', and as a matter of fact, implies that customers who sense in Ω also sense in Ω'.

We assign each customer one of three types based on the values of their sensing aversion: Customer i will be assigned

- *type-L* (stands for “Low”) if $u_i \in [0,p]$,

- *type-M* (stands for "Medium") if $u_i \in (p, p']$,
- *type-H* (stands for "High") if $u_i \in (p', 1]$.

By definition, upon arrival, an H-type customer joins both S_Q and S'_Q, the shared queues in Ω and Ω' respectively, an M-type customer joins S_Q but senses S'_L, whereas an L-type customer senses both S_L and S'_L.

W.l.o.g., we modify the service regime as follows:

(i) Arriving at Ω (Ω') when S_L (S'_L) is busy, customer i preempts the customer in service in S_L (S'_L) with no regard to type, and the preempted customer is routed to continue the service in S_Q (S'_Q). Arriving at Ω (Ω') when S_L (S'_L) is empty, customer-i's actions follow by type.

(ii) The shared queue, subsystem S_Q (S'_Q), is a preemptive resume LCFS-PR queue, meaning that customers are served following a preemptive last-come-first-served discipline, and the preempted service resumes from the last point it stopped.

Lemma 8.1.1. *Assumptions* (i) *and* (ii) *do not affect the stationary probabilities of the systems.*

Proof. We shall show that Lemma 8.1.1 holds for system Ω and the proof for Ω' is identical.

Let i and $i+1$ be a successive pair of customers and suppose that customer $i+1$ preempts customer i in subsystem S_L. Denote $\Delta_{i,i+1} := T_{i+1} - T_i$ the time difference between their arrivals. Then, as a result of the memoryless property of the exponential distribution, the service duration of customer $i+1$ (i.e., τ_{i+1}) and the residual service of customer i (i.e., $\tau_i - \Delta_{i,i+1} \mid \tau_i > \Delta_{i,i+1}$) are independent and identically distributed. This incident of preemption is equivalent to the joining of the new comer to subsystem S_Q, the shared queue, when S_L is busy. Hence, assumption (i) does not influence the transition probabilities and the stationary probabilities remain as before.

The validity of assumption (ii) can be explained by the fact that in a work-conserving preemptive resume system with exponential service duration, as sustains in S_Q, the queue length is independent of the service regime. Consequently, assumptions (i) and (ii) do not restrict the generality of the model. $\square$

Exploiting assumptions (i) and (ii) we immediately achieve several properties that will be later used in the proof:

Lemma 8.1.2. *Under assumptions* (i) *and* (ii)*, the following arguments hold:*

(a) *At any moment, if S_L (S'_L) is busy, then the last customer to arrive until that moment is in S_L (S'_L).*

(b) *Both in Ω and in Ω', all customers begin service at the moment of arrival, whether they sense or not.*

(c) *Customers beginning service in S_L upon arrival, begin service in S'_L as well.*

(d) *Customers beginning service in S'_Q upon arrival, begin service in S_Q as well.*

(e) *The sojourn time of customer i in S'_Q is no longer than the sojourn time of i in S_Q.*

Proof.

(a) Suppose i is the last customer until time t, and i is not in S_L at time t. From assumption (i), this is possible if either i is of type L and S_L is idle at the moment of i's arrival, or if i joins S_L and completes service before time t. Hence, in both scenarios, S_L stays idle at time t. The proof for Ω' is similar.

(b) In system Ω, customers at their moment of arrival either begin service immediately in S_L, or, by assumption (ii), join the head of the shared queue, thus beginning service in S_Q. The proof for Ω' is similar.

(c) Recall that at time $t = 0$ both S_L and S'_L are empty. Only upon the arrival of an L-type customer will the state of S_L alter. Arriving at Ω', an L-type customer joins S'_L, regardless of the server's state. Suppose the last customer has joined both S_L and S'_L. By assumption (i), this tagged customer can either complete service in S_L and S'_L simultaneously (so S_L and S'_L become empty) or be preempted in the two subsystems by the next customer to come. As for the latter case, it holds that the preempting customer joins both S_L and S'_L. Thus, it follows by induction that customers who join S_L also join S'_L.

(d) This property is immediately achieved as the *modus tollens* form of property (c).

(e) It results from assumption (ii) that the sojourn time of each customer in the shared queue depends only on future arrivals. Combining this with property (d) we deduce (e).

□

With property (c) and the simultaneousness of occurrences we get:

$$\forall t \in [0, \infty) : \{Y(t) = 1\} \Rightarrow \{Y'(t) = 1\} ,$$

or, in an analogous fashion:

$$\forall t \in [0, \infty) : \{Y'(t) = 0\} \Rightarrow \{Y(t) = 0\} . \tag{8.3}$$

Moreover, from properties (d) and (e) we see that, if at an arbitrary point in time there is a customer waiting in S'_Q, then the same customer will also be waiting in S_Q, not leaving Ω before leaving Ω'. Therefore,

$$\forall t \in [0, \infty) : X'(t) \leq X(t) , \tag{8.4}$$

implying that the process representing the number of customers in S_Q over time stochastically dominates the one representing the number of customers in S'_Q.

For the sake of simplifying the notation, when there is no ambiguity we denote $(X(t), Y(t))$ and $(X'(t), Y'(t))$ by (X, Y) and by (X', Y') respectively. On utilizing (8.4), we get

$$\mathrm{E}[X' \mid Y' = 0] \leq \mathrm{E}[X \mid Y' = 0] .$$

To prove our claim (8.1) it suffices to prove:

$$\mathrm{E}[X \mid Y' = 0] \leq \mathrm{E}[X \mid Y = 0] . \tag{8.5}$$

Denote by i the last customer who arrived by time t, and let

$$A(t) := \{i \text{ joined } S_Q \text{ and } T_i + \tau_i > t\}$$

express the event "i joined the queue in Ω and was still in the system by time t". For i being the last customer, by property (b) we get that $A(t)$ occurred only if i was still in service in S_Q by time t. Therefore from property (a),

$$\forall t \in [0, \infty) : A(t) \Rightarrow \{Y(t) = 0\} . \tag{8.6}$$

Recall that by property (a) if there is a customer in S'_L at time t then it must be i (the last customer arrived). So, assuming that $Y(t) = 0$ and $Y'(t) = 1$, we have that customer i, arriving at Ω, joined S_Q, and at the same time, arriving at Ω', began service in S'_L. As i was the last customer until time t, knowing that $Y'(t) = 1$ implies that $T_i + \tau_i > t$. Hence,

$$\forall t \in [0, \infty) : \{\{Y'(t) = 1\} \cap \{Y(t) = 0\}\} \Rightarrow A(t) . \tag{8.7}$$

Lemma 8.1.3. *Conditioned on $A(t)$, $X(t)$ and $Y'(t)$ are independent random variables.*

Proof. Let t be an arbitrary point in time and i be the last customer to arrive by that time. Assuming that $A(t)$ had happen, it means that i was still in Ω by time t. The residual service of customer i at time t is an exponential variable with parameter μ. Thus, we can assume w.l.o.g. that i's arrival (at time T_i) occurred a short moment before t. $A(t)$ can be rewritten then as the event when a new busy period in S_Q began right before t (i.e., $t = T_i^+ := T_i + \varepsilon$, where ε is a small positive infinitesimal quantity).

Denote T_n^+ the moment after customer n's arrival. To prove Lemma 8.1.3 we shall show by induction that $\forall n \in \mathbb{N}$, $X(T_n^+)$ and $Y'(T_n^+)$ are conditionally independent given $A(T_n^+)$.

Presume that $A(T_n^+)$ is satisfied. Having that n joined S_Q means that n is not of type L, and, by (8.6), $Y(T_n^+) = 0$. Define $T_n^- := T_n - \varepsilon$, where ε is a small positive infinitesimal quantity, the moment before n's arrival. Then,

$$X(T_n^+) | A(T_n^+) = X(T_n^-) + 1 . \tag{8.8}$$

Regarding n's arrival at Ω', there are two possibilities:

- Upon n's arrival, S'_L was empty (i.e., $Y'(T_n^-) = 0$).
- Upon n's arrival, S'_L was busy (i.e., $Y'(T_n^-) = 1$).

To the extent that $A(T_n^+)$ holds and $Y'(T_n^-) = 0$, n's action (and therefore the state of S'_L) is determined one to one by n's type (M or H):

$$Y'(T_n^+) \mid A(T_n^+) = \begin{cases} 0, \text{ iff } n \text{ is of type-H,} \\ 1, \text{ iff } n \text{ is of type-M;} \end{cases} = \begin{cases} 0 \text{ w.p. } \frac{1-p'}{1-p}, \\ 1 \text{ w.p. } \frac{p'-p}{1-p}. \end{cases}$$

Since i's type does not depend on previous events in the system it follows from (8.8) that $X(T_n^+)$ and $Y'(T_n^+)$ are independent conditioned on $A(T_n^+)$.

To the extent that $A(T_n^+)$ holds and $Y'(T_n^-) = 1$, with assumption (i), n must have preempted his/her predecessor, $n - 1$, in S'_L, simultaneously as joining S_Q. This is possible only if customer $n - 1$ remained in service until customer n's arrival, and indicates that

(a) Customer $n-1$ joined S_Q upon arriving at Ω, and $Y(T_{n-1}^+) = 0$.

(b) Customer $n-1$ joined S'_L upon arriving at Ω', and $Y'(T_n^+)|A(T_n^+) = Y'(T_{n-1}^+) = 1$.

(c) Customer n had arrived before $n-1$ left S_Q, and $X(T_n^+)|A(T_n^+) = X(T_{n-1}^+) + 1$.

From (a), (b) and (8.7) we deduce that $A(T_{n-1}^+)$ holds. Followed by the induction hypothesis, $X(T_{n-1}^+)$ and $Y'(T_{n-1}^+)$ are independent conditioned on $A(T_{n-1}^+)$. Thus, from (b) and (c) we conclude that $X(T_n^+)$ and $Y'(T_n^+)$ we are also independent conditioned on $A(T_{n-1}^+)$.

It is left to show that $X(T_1^+)$ and $Y'(T_1^+)$ are conditionally independent given $A(T_1^+)$. We can assume, w.l.o.g., that $Y'(T_1) = 0$ (otherwise we observe the last customer that arrived before T_1). Thus, as discussed, cusotmer 1's action is equivalent to type, which is a random variable independent of $X(T_1^+)$. □

Define the operator $\mathrm{E}_{Y=0}[\bullet] := \mathrm{E}[\bullet \mid Y = 0]$. Combining Lemma 8.1.3 with (8.7) we arrive at

$$\mathrm{E}_{Y=0}[X \mid A] = \mathrm{E}_{Y=0}[X \mid A, Y' = 1] = \mathrm{E}_{Y=0}[X \mid Y' = 1] \,. \tag{8.9}$$

Since $A(t)$ indicates the beginning of a new busy period at time t

$$\mathrm{E}_{Y=0}[X] \leq \mathrm{E}_{Y=0}[X \mid A] \,.$$

Upon substituting (8.9) we get

$$\mathrm{E}_{Y=0}[X] \leq \mathrm{E}_{Y=0}[X \mid Y' = 1] \,. \tag{8.10}$$

By the law of total expectation we have

$$\begin{aligned} \mathrm{E}_{Y=0}[X] &= \mathrm{E}_{Y=0}[X \mid Y' = 0] \cdot \Pr(Y' = 0 \mid Y = 0) + \\ &\quad\ \mathrm{E}_{Y=0}[X \mid Y' = 1] \cdot \Pr(Y' = 1 \mid Y = 0) \\ &= \mathrm{E}[X \mid Y' = 0] \cdot \Pr(Y' = 0 \mid Y = 0) + \\ &\quad\ \mathrm{E}_{Y=0}[X \mid Y' = 1] \cdot \Pr(Y' = 1 \mid Y = 0) \,, \end{aligned} \tag{8.11}$$

where the second equality follows from (8.3). Note that the right-hand side of (8.11) is a convex combination of $\mathrm{E}[X \mid Y' = 0]$ and $\mathrm{E}_{Y=0}[X \mid Y' = 1]$. Amalgamating (8.10) and (8.11) we attain

$$\mathrm{E}[X \mid Y' = 0] \leq \mathrm{E}_{Y=0}[X] \,,$$

thus confirming (8.5). This completes the proof of Proposition 4.1. □

8.2 Proof of Proposition 5.1

Proof. Given that $c_s = 0$ and ${}^{c_w}\!/\!_{\mu} = 1$, the expected cost of a customer ($F(p)$) is the expected number of customers ahead of him/her, which varies depending on his/her action (sensing or not sensing). For some δ such that $p + \delta \leq 1$, define $D(p)$ as the proportionate difference in social cost, on average, incurred by increasing the proportion of sensing customers in the system by δ, namely,

$$\delta \cdot D(p) = F(p + \delta) - F(p) \,. \tag{8.12}$$

Lemma 8.2.1. *$D(p)$ is monotone non-decreasing in p.*

Proof. As in the proof of Proposition 4.1, recall $\Omega = \{S_Q, S_L\}$ and $\Omega' = \{S'_Q, S'_L\}$ with the states $(X(t), Y(t))$ and $(X'(t), Y'(t))$ and sensing probabilities p and p' ($p < p' < 1$) respectively. We use $\{(T_i, \tau_i, u_i)\}_{i \in \mathbb{N}}$ - the set of variables defining the queueing process as in (8.2), with T_i the arrival time, τ_i the service duration and u_i the sensing aversion of customer i. We also assume (i) and (ii) as in Lemma 8.1.1.

Tag customer i such that $u_i > p'$. By definition, customer i does not sense, neither in Ω nor in Ω'. Thus, forcing this tagged customer into sensing strictly increases the population of sensing customers within the two systems. In addition, the fact that each customer that senses in Ω also senses in Ω' is still valid, weather if we say that i senses in both systems or if we say that i does not sense in any system. This ensures that Lemma 8.1.2 holds in each of the two cases, i senses and i does not. We intend to investigate each system, Ω and Ω', under each of the two circumstances: When customer i is forced to sense and when he/she does not sense. Define:

- D_i - The difference between the social cost incurred by i when he/she senses and when he/she does not sense in system Ω.
- D'_i - The difference between the social cost incurred by i when he/she senses and when he/she does not sense in system Ω'.

Recall by (8.3) that for all $t \in [0, \infty)$, $\{Y'(t) = 0\} \Rightarrow \{Y(t) = 0\}$, thus,

$$\forall t \in [0, \infty) : \Pr(Y(t) = 1, Y'(t) = 0) = 0 \,.$$

We divide our event space, therefore, into three disjoint cases: When $Y(T_i) = 1$, when both $Y(T_i) = 0$ and $Y'(T_i) = 1$, and when $Y'(T_i) = 0$.

(a) If $Y(T_i) = 1$ (consequently, by (8.3), $Y'(T_i) = 1$) then, with the fact that $c_s = 0$, the social planer in Ω is indifferent between i sensing and i not sensing as both actions result in the same outcome, thus $D_i = 0$. In a similar manner, since $Y'(T_i) = 1$ in Ω', $D'_i = 0$.

(b) If $Y'(T_i) = 1$ and $Y(T_i) = 0$, as explained above, $D'_i = 0$. Regarding system Ω, when customer i senses there is the possibility that i will complete service in S_L and $X(T_i) + 1$ customers (that is $X(T_i$ customers in the queue plus customer i) would save the expense incurred by i joining the queue upon arrival. Thus the social cost when i senses is less than when he/she does not sense and $D_i \leq 0$.

(c) If $Y'(T_i) = 0$ (consequently, by (8.3), $Y(T_i) = 0$) then we separate between two cases:

  - When $\tau_i \leq T_{i+1} - T_i$, customer i completes service in S_L (and in S'_L) before the arrival of the next customer. Denote by B_i and B'_i the sojourn time of i in the queue of Ω and of Ω' respectively, measured in units of $1/\mu$. From property (e) of Lemma 8.1.2 it is clear that $B'_i \leq B_i$. Since $c_w/\mu = 1$, i sensing in Ω in this case saves $X(T_i)B_i + B_i - 1$ relative to not sensing (that is $X(T_i)$ customers in the system saving B_i each, plus customer i saving B_i minus the cost of his/her own service). Likewise, i sensing in Ω' saves $X'(T_i)B'_i + B'_i - 1$. Recall by (8.4) that $X'(T_i) \leq X(T_i)$, then we have:

$$D_i = 1 - B_i - X(T_i)B_i \leq 1 - B'_i - X'(T_i)B'_i = D'_i \,.$$

- When $\tau_i > T_{i+1} - T_i$, customer i, when sensing, begins service in S_L (S'_L) but with assumption (i) he/she is forced to join the queue upon the arrival of $i+1$. Denote by L and L' the number of customers that has left the system during the time period $[T_i, T_{i+1})$ in Ω and in Ω' respectively. Since $X'(T_i) \leq X(T_i)$, $L' \leq L$. Denote by B_i and B'_i the sojourn time of i in the queue of Ω and of Ω' respectively, measured in units of $1/\mu$. As explained, $B'_i \leq B_i$. Then we have L customers in Ω saving the expense of waiting B_i each one and L' customers in Ω' saving the expense of waiting B'_i each one, therefore,

$$D_i = -LB_i \leq -L'B'_i = D'_i \, .$$

In summary, for i such that $u_i > p'$, under all circumstances $D_i \leq D'_i$. Subsequently, if $\mathcal{S} \subset \mathbb{N}$ is a subset of customers such that for all $j \in \mathcal{S}$, $u_j > p'$, then

$$\frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} D_j \leq \frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} D'_j \, .$$

This implies that for a given δ, when increasing p by δ, the marginal social cost gets bigger with the increase of p, i.e.,

$$\delta D(p) = F(p+\delta) - F(p) \leq F(p'+\delta) - F(p') = \delta D(p') \, ,$$

where the first equality and the last equality follow from (8.12) and we conclude that $D(p)$ is monotone non-decreasing. $\square$

Note from (8.12) that $D(p) = 1/\delta (F(p+\delta) - F(p))$, so the monotonicity of $D(p)$ implies that $(d/dp)F(p)$ is non-decreasing, and therefore $F(p)$ is convex in p. $\square$

References

[1] Chuong T. Do, Nguyen H. Tran, Mui Van Nguyen, Choong Seon Hong, and Sungwon Lee. Social optimization strategy in unobserved queueing systems in cognitive radio networks. *IEEE Communications Letters*, 16:1944–1947, 2012.

[2] Noel M. Edelson and David K. Hildebrand. Congestion tolls for poisson queueing processes. *Econometrica*, 43:81–92, 1975.

[3] Oussama Habachi and Yezekael Hayel. Optimal opportunistic sensing in cognitive radio networks. *IET Communications*, 6:797–804, 2012.

[4] Refael Hassin. *Rational Queueing.* CRC Press, Boca Raton, 2016.

[5] Refael Hassin and Moshe Haviv. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems.* Kluwer Academic Publishers, Boston, 2003.

[6] Moshe Haviv and Ya'acov Ritov. Externalities, tangible externalities, and queue disciplines. *Management Science*, 44:850–858, 1998.

[7] Yezekael Hayel, Dominique Quadri, Tania Jimnez, and Luce Brotcorne. Decentralized optimization of last-mile delivery services with non-cooperative bounded rational customers. *Annals of Operations Research*, 239:451469, 2014.

[8] Simon Haykin. Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23:201–220, 2005.

[9] Krishna Jagannathan, Ishai Menache, Eytan Modiano, and Gil Zussman. Non-cooperative spectrum access - the dedicated vs. free spectrum choice. *IEEE Journal on Selected Areas in Communications*, 30:2251–2261, 2012.

[10] Husheng Li and Zhu Han. Socially optimal queuing control in cognitive radio networks subject to service interruptions: To queue or not to queue? *IEEE Transactions on Wireless Communications*, 10:1656–1666, 2011.

[11] Stephen C. Littlechild. Optimal arrival rate in a simple queueing system. *International Journal of Production Research*, 12,:391–397, 1974.

[12] Uri Yechiali and Pinhas Naor. Queuing problems with heterogeneous arrivals and service. *Operations Research*, 19:722–734, 1971.